DOCUMENT RESUME

ED 178 103 IR 007 893

AUTHOR Caruso, Elaine
TITLE Online Training for Searching Online.
PUB DATE Dec 78
NOTE 14p.; Paper presented at the International Online Information Meeting (2nd, London, December 5-7, 1978); For related documents, see IR 007 891-892

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Autoinstructional Programs; Bibliographic Citations; *Computer Assisted Instruction; Data Bases; *Information Retrieval; *On Line Systems; *Program Development; Questioning Techniques; *Search Strategies; Use Studies

ABSTRACT

This paper focuses on the casual user's needs and problems in learning to search online retrieval systems. Such users are inclined to infrequent use, and are unlikely to have much interest or knowledge about the complexities of the system. The TRAINER program is designed to teach these users to search online, so that they may access information files without the aid of an intermediary. Three user problems are earmarked: (1) inability to perceive characteristics of different data files, (2) decisions regulating structure and coordination of document file indexes, and (3) search language attributes. To develop the TRAINER program, the online search process and decision-making complexity were analyzed from the user's perspective. Two kinds of programs comprise the TRAINER system: computer assisted learning modules, and practice modules (system emulators). Users seem to prefer exercises prepared for use with the system emulators for refinement of search skills. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED178103

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

2nd International Online Information Meeting

London 5-7 December 1978

Organised by
Online Review
THE INTERNATIONAL
JOURNAL OF ONLINE &
TELETEXT INFORMATION
SYSTEMS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Elaine Caruso

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR007893

Learned Information Oxford and New York

PAPER B5

ONLINE TRAINING FOR SEARCHING ONLINE

Elaine Caruso, University of Pittsburgh, U.S.A.

Keywords: Online Information Systems, Computer Assisted Learning, DIALOG, ORBIT, End-User Training Emulators of Online Search Services.
Abstract: Training users for online bibliographic searching is difficult. Problems arising from multiple sources of data files, index building practices, and deficiencies of the search algorithms are compounded when trainees must learn to use a new tool, the search system, and a new process, the online search, simultaneously. TRAINER, using computer assisted learning modules and emulators of the DIALOG and ORBIT systems, provides a responsive and protected environment in which users may develop skills in use of the tool and in conducting the search process.

1 INTRODUCTION

Training a new user to operate an online bibliographic retrieval system is difficult to do. Each line entered from the user's terminal requires more decisions than are comfortable, especially when seconds of delay are costing an appreciable amount of money for communication charges and for computer-connect time.

As a consequence, most online searchers are individuals who accept the challenge of this demanding environment as a part of their particular professional repertoire. These individuals learn to function in a mediating capacity between an individual who has an information need and the storage and retrieval system with its multiple data files.

According to circumstantial variables, this may be a desirable or undesirable mode of using the search service. Where an information specialist is available, it is likely that he will be more competent than the researcher himself in optimizing actual cash costs of the search, in achieving desired specificity or comprehensiveness of search results, and in coordinating research needs when several individuals are cooperating on a common activity. The intelligence, knowledge of the users' needs, online training and experience and the immediate motivation of the search intermediary may contribute substantially to the usefulness of the retrieval service.

On the other hand, the user's real need to browse, however inefficiently, can be completely forestalled in the machine readable data file era if we allow dependency on this intermediary to become the rule. Economic and intellectual barriers militate to prevent our achievement of universal access to machine readable data files (MRDF) which the Information Industry proclaims as a present possibility. (Ref. 1) Whereas current technology would permit such wide access, if it is predicated on the

concurrent availability of trained human interfacers, it is not credible. The cost of such service would negate the extension of access which improvements in the technology have made possible.

Our profession has an obligation to make direct user searching feasible, both economically and intellectually. Improvements in the retrieval system, and appropriate, accessible user training are aspects of this obligation.

2 TRAINER: PREMISES

The Trainer system has been developed to meet the demand for training suitable to what has been called the casual user, as opposed to a specialist in system use. (Ref. 2) The casual user is (1) inclined to infrequent use, and (2) is unlikely to have much interest in the internal workings of the system, or (3) to have acquired the full system documentation or the many vocabulary aids or detailed file descriptions. In Trainer we assume he will have minimal system documentation and only the briefest file descriptions as provided by the system(s) in use. (Ref. 3, 4)

The Trainer has been documented elsewhere; it is designed for use in the same mode, using the same equipment as the online services themselves. (Ref. 5, 6) My purpose here is to tell you what we have learned about the services themselves, and the user, in our development of the modules which first assist the user in learning about the services (Computer Assisted Learning Modules) and second, allow him to practice the skills to develop facility in their use (the Emulator Modules).

Observation of users in training suggests that printed documentation is used reluctantly, if at all. The Trainer itself relies minimally on paper texts, and we are, furthermore, convinced that effective use of the data files themselves can be made without recourse to the printed thesauri and vocabulary lists which are recommended as essential in search preparation. The brief one or two page file descriptions provided by the search service will suffice; the more complete descriptions provided by file developers for distribution by the search service are very helpful <u>if</u> users will take the time and effort to study them.

Our work with the trainer program has demonstrated that the minimal support documents can suffice for the development of a search. A searcher can use his own vocabulary to get several document records printed online. File development by the service always includes some "natural language" fields in the indexes, in addition to the more formal indexing provided. Title words and added keywords are indexed so that the user's initial natural language search vocabulary is likely to find "hits" even though the original file used only a very rigidly controlled indexing vocabulary. These "hits" provide clues to controlled vocabulary search terms which can then be used to retrieve additional documents in the file.

3 PROBLEMS OF THE ONLINE SEARCH

Why then is online searching perceived to be so very difficult to do well, even for those who have more than casual acquaintance with the systems and data files? It is that the user, working from the outside, seeing only the surface of the machine, cannot perceive what is going on inside. The fact that there are differences in the separate files, which would matter little if you could see the printed pages, is one source of difficulty. Figure 1 shows how names can vary from one data file to another. In today's practice this variation can result in retrieval failure, as shown in Figure 2.

Another kind of problem is caused by decisions of the search services about their structuring of the indexes to the document files, and the way the separate indexes to document record fields are combined for searching as "basic" or "subject" indexes, or as separate indexes available only when the system designated search term qualifiers are correctly used. Figure 3 illustrates a failure caused by an incorrect term qualifier; Figure 4 shows how variation in policies for handling multiword index terms requires users to vary the search terms they enter.

A third kind of problem results from search language attributes; either from the unsophisticated (that is, unforgiving) processing which interprets user commands, or the rigidity of the algorithms for matching user search terms with file indexes. As the distinction between upper and lower case is ignored, the algorithms could as well equate variations in spacing, punctuation, hyphenation, and even spelling in determining a "match."

Responsibility for easing these difficulties is not easily assignable (See Figure 5). Changes either in data file practices of file suppliers, or in data base building, by the retrieval system, or in the search languages, used in the search system, could relieve the problems; but with such diffusion of possible solutions it is not likely that any one of the identified search system components will initiate needed changes. In practice it devolves on the fourth component of the document retrieval system, the searcher, to compensate for system inadequacies; and in articulating his difficulties, to provide feedback about them to the accessible, responsive service or data file suppliers' representatives.

4 COMPLEXITY IN THE ONLINE PROCESS

In our development of the Trainer programs, we began by analyzing the online search process in several ways, but always from the outside, the user's perspective. One such analysis is reflected in our tutorial modules (Figure 6); we separated out operations which could be used optionally, and only where needed: e.g., use of the present host system (University of Pittsburgh) login procedures, and use of TYMNET and TELENET. We provided an overview, in Module III, of the complete process, and then, in other separately accessible modules, IV-VII, the several procedures which could be readily isolated and dealt with.

Another analysis resulted in our selection of the specific commands and capabilities to be taught (Figure 7). Even this simplification did not reduce the complexity of the task to be learned; a line by line analysis of the task revealed that a beginning searcher needed to keep six or more decision factors in mind in entering the simplest command. This does not include the processes involved in verbalizing the question to be answered or selecting the file(s) to be searched. While question formulation and file choice are preliminary to the online search, they are not completed steps but continuing aspects of the online search. They affect each step of the search and are themselves quite likely to be changed as the search proceeds.

Consider just the first activity; specifying to the system which file is to be used:

1. What command is used:
   - if more than one command is available,
   - what are they?
   - which one is needed - for economy? for special capabilities?

2. Which form of command?
   full form - for ease and speed of recall? for readability?
   abbreviation - for economy of keying? to eliminate typing errors? - if you remember the exact abbreviation?
   symbolic form - if no one character abbreviation is available? (if you can remember it)?
3. How is file specified - by abbreviated or specific form of name? or by assigned number?
4. What is the abbreviation or number of the desired file?
5. In entering the command sentence, should one eliminate spacing to save time? or is that so foreign to user habits that it causes him to slow down? or, perhaps a space is required between the name of the command and the file name or number--or is it required that there be no space?
6. If one makes a typo should one try to correct it? how? or would it be quicker/more economical to allow the system to process and reject it, and reenter it?

Some of the questions are pertinent at every line of input; other decisions to be made become more complex as the search proceeds, and the searcher is required to incorporate what he has learned from previous commands, to consider more complex options, and to anticipate the effects c file size, postings density, and file organization. For example, what is in the default index when field qualifiers are not used? What are the field labels?

Our object in making this particular kind of analysis was to gain further insight into the nature of the complexity; to make logical clumps of associated procedures; to strip down further the set of basic commands to be taught. Although we became aware that the necessity for making many of those decisions could be ascribed to practices in one of three distinct responsible subsystems of the total system, and that many of these can be and should be eliminated since they do not inhere to the online search nor contribute to its value or effectiveness, we still recognize that users of present systems must cope.

5 OUR APPROACH TO TRAINING

Our first effort was devoted to the identification of a useful but simplified subset of the online environment. The next was to create the online programs to enable the beginner to learn how to operate in that environment. We built two kinds of programs: the computer assisted learning (CAL) modules and the practice modules, the system emulators. In both cases the attempt was to mimic the real world system; but in the CAL modules some direction for proceeding is given initially, and if the user makes errors, explanations and assistance are given. In the practice modules the user gets only the error responses he would have in actual use of the systems themselves. (Ref. 7)

An extensive Manual for users is available for those who wish to use it; our observation is that, in fact, it is used very little as a learning aid, but serves as a reference book for later examination of specific commands or processes. That is, users seem to prefer to learn by hands-on trial and error.

The most useful printed aid seems to be, by user perception, that set of exercises which we prepared to guide the trainee in practice searches on the system emulators, to flesh out the notions of search statement

creation and refinement and output control, which the CAL modules introduce. As the trainee uses the emulators he learns to relegate the problems of command language and syntax to a background level of processing, and to focus his conscious thinking on better expression of the question, and on interpretation of system feedback for more effective exploitation of the particular file being searched.

6 CONCLUSION

To the outside world, information science is most visible in the information system developments which make the use of machine readable data files possible. Users of those systems are beginning to make known their dissatisfactions. Dr. William Carter of Harry Diamond Laboratories said very clearly, in commenting on an Applications Working Group report at the NATO Institute, "your theory is fine, but your application is deficient; promise is great, but practice is poor." (Ref. 8)

As a group we need to bring our attention to the improvement of the systems being developed now, and of those **already in wide use.** Also I believe that efforts to develop a larger, more diversified population of users, online searchers who will have no vested interest in maintaining the status quo, is not only justified, but mandated. It is justified in that it will require more usable systems to be developed; it is mandated by our professional responsibility to facilitate the use of knowledge.

REFERENCES
1. Paul Zurkowski, President, Information Industry Association. Paper presented to NATO Advanced Study Institute in Information Science, Chania, Crete, August 1978.
2. U.S. National Science Foundation Grant No. DSI 77-26525, Training Modules for Users of Scientific and Technical Information Services.
3. A Brief Guide to DIALOG* Searching. Lockheed Information Systems, Palo Alto, California, 1976.
4. ORBIT Quick-Reference Guide. System Development Corp., Santa Monica, California, 1977.
5. Training Modules for Use of Scientific and Technical Information Services. Report, June 1, 1976-November 30, 1977, Dec. 1977.
6. _____. Work in Progress, Mar. 1, 1978-May 31, '78, June, 1978.
7. Ibid. Appendix 2, Transcripts which show how Trainer's CAL and Emulators mimic the 'real' system, until the user makes an error or needs help.
8. William Carter, Technical Director, Harry Diamond Laboratories. Report of Technical Applications Working Group, NATO Advanced Study Institute in Information Science, Chania, Crete, August 1978.

PREPARING TO RUN A NAME SEARCH:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
BIOSIS PREVIEWS	C	A	R	U	S	O			D			E					
CHEMCON	C	A	R	U	S	O	,			D	.	E	.				
ENVIROLINE	C	A	R	U	S	O	,			D	O	R	O	T	H	Y	E
GEO REF	C	A	R	U	S	O				D	.		E	.			
NTIS	C	A	R	U	S	O	,			D	O	R	O	T	H	Y	E
SSIE	C	A	R	U	S	O			D	E							
	C	A	R	U	S	O			E	L	A	I	N	E			
	C	A	R	U	S	O				E	.						
	C	A	R	U	S	O	,			E	L	A	I	N	E		

Fig. 1. Different practices for entering personal names in machine readable data files are illustrated. These forms are not normalized by search services in creating the searchable indexes. Consistency of punctuation, alone, would greatly simplify the 'matching' process.

A SEARCH PROGRAM VARIATION TO COPE WITH NAMES:

TRUNCATED NAME FORMS

1. S AU = CARUSO?

   CREATES A HIT FILE OF 41 ITEMS (NTIS)

=OR=

2. S AU = CARUSO, D?

   CREATES A HIT FILE OF 2 ITEMS (NTIS)

PROBLEMS:

EX. 1 ALL OF THE 41 ITEMS ARE NOT WANTED:

EX. 2 WE WILL MISS ITEMS LISTED:

CARUSO D E
CARUSO DE
CARUSO E
CARUSO, ELAINE

Fig. 2. In order to cope with various practices a searcher must truncate the name form immediately after the personal surname. This produces unmanageable output, even with relatively uncommon names. The next possible truncation, following first initial, misses all forms entered without punctuation. Library practice, requiring research to establish a 'name authority file' is necessary to circumvent problems caused by authors who use various 'pen names'; as illustrated here, the use or non-use of various 'given' names.

NOT ALL PROBLEMS OF PERSONAL NAME SEARCHING ARE SOLVED BY TRUNCATION:

HERE IS ONE SEARCH RUN ON THE SSIE FILE, ON TWO DIFFERENT SEARCH SYSTEMS:

FIRST SYSTEM:	
S AU=CARUSO, D	NO HITS
S AU=CARUSO?	NO HITS
SECOND SYSTEM:	
CARUSO, D:/AU	NO HITS
CARUSO:/AU	6 VARIANTS:
	1. CARUSO DE/IN
	2. CARUSO FM/IN
	ETC.

Fig. 3. The "first system" requires the searcher to use the search term qualifier IN (INVESTIGATOR) to locate the author of a research proposal. The "second system" recognizes either IN or AU. Neither system helps the user who fails to notice that the SSIE file does not punctuate personal name entries.

NTIS

System 1: S AU=CARUSO, D?

System 2: CARUSO, D:/AU

System 1: INFORMATION RETRIEVAL
AND TRAINING
AND ON AND LINE

System 2: INFORMATION
AND
RETRIEVAL
AND ON AND LINE

Fig. 4. The author search on the NTIS file is identical in the two systems. "System 2" made a decision to split the NTIS supplied subject index terms into single words, believing that the often long phrases would be less useful as retrieval keys. "System 1" uses the terms as supplied, and also as single word index entries.

WHO IS RESPONSIBLE?

1. THOSE WHO CREATE FILES?
2. THOSE WHO DESIGN, STRUCTURE THE SEARCHABLE FILES FOR ONLINE SEARCH?
3. THE SEARCH LANGUAGE DESIGNER?
4. THE SEARCHER?

Fig. 5. Problems arise from errors in keying entries to the MRDF and from inconsistent practices within and between different files. Practices of the different retrieval services vary from each other, and from file to file within their own data bases. Search languages are not adequately designed to compensate for file errors and inconsistencies. Any of these three subsystems of the retrieval system could be expected to correct the situation, but none do. Therefore the user, at time of use, must compensate.

Modules I-VII Computer Assisted Learning

I. Login to Host Computer System
II. Using TYMNET and TELENET to Access LRS and/or SDC
III. Creating a Complete Search, DIALOG and/or ORBIT
IV. Choosing and Using a Data File
V. Negotiating the Search Terms DIALOG, ORBIT, and/or SCORPIO
VI. Using Logical Operators in Search Statements, DIALOG and/or ORBIT
VII. Formatting and Printing Selected Items

Modules VIII-IX Practice in Systems Emulators

VIII. DIALOG Emulation, 9 file subsets
IX. ORBIT Emulation, 9 file subsets

Fig. 6. Trainees may enter the program at any of the identified modules, and within the module for any identified system; for example, the usual starting point is Module III, DIALOG. Average use time for a given system in a single module: 5 minutes. Searches in the emulators tend to be equally brief, about 7 minutes per search.

TO REACH TRAINER ONLINE
****
DIAL (412) 621-5954
↑C ←
SET TTY SYS A ←
LOG 134057,_ _ _ _ _ _ ← (1)
PASSWORD→ _ _ _ _ _ _ ← (2)

FOR CALF MODULES:
****
R CTLYST ←
→STI.001[134057,120121]←

FOR DIALOG EMULATOR:
****
RUN DIALOG[134057,120121] ←

FOR ORBIT EMULATOR:
****
RUN ORBIT[134057,120121]←

TO REACH LRS ONLINE:
****
TYMNET: TELENET:
DIAL_ _ _ _ _ _ (3) DIAL _ _ _ _ _ (3)
E ← ←
LRS ← ←
DIALOG ← C 415 20 ←
PASSWORD ← (2) PASSWORD ← (2)

TO REACH SDC ONLINE:
****
TYMNET: TELENET:
DIAL_ _ _ _ _ _ (3) DIAL _ _ _ _ _ (3)
E ← ←
SDC ← ←
ORBIT ← C 213 33 ←
/LOGIN USERID←/LOGIN USERID← (2)

TRAINER DATA BASE JAN. '78
****
1 OCEANIC ABSTRACTS
2 INSPEC
3 ENGINEERING INDEX
4 POLLUTION ABSTRACTS
5 PAPER CHEMISTRY
6 GEOREF
7 BIOSIS
8 NTIS
9 ERIC

****
← INDICATES CARRIAGE RETURN
(1) USE YOUR ASSIGNED NUMBER
(2) USE YOUR ASSIGNED PASSWORD
****
(3) Use local phone no.

TRAINER COMMANDS
****
HELP
EXIT
E/F (logs off Pitt)

CHARACTER CANCEL← ↑H
LINE CANCEL← ↑U
INTERRUPT← ↑O

ORBIT COMMANDS
****
FILES?
FILE _ _ _ FILE ERIC
FIND _ _ _ FD term
NBR _ _ _ NBR term
(TERM ENTRY) term
(LOGICAL COMB.) term OR term
term AND term
term AND NOT term
PRINT PRT TRIAL SS 2
PRT 6 AU, TI
HISTORY HIS HIS REVERSE
STOP
CHARACTER CANCEL←↑H (varies)
LINE CANCEL← $←
INTERRUPT← BREAK←

DIALOG COMMANDS
****
? EXPLAIN ?FILES
! BEGIN n ↑6
.FILE .FILE6
EXPAND E term
PAGE P
SELECT S term
COMBINE C1+2 C1*2 C1-2
TYPE T6/2/1-5
& PRINT PR6/2/1-5
&-PRINT- PR-
= END
DISPLAY SETS DS
LOGOFF

CHARACTER CANCEL← ↑H
LINE CANCEL← BREAK←
INTERRUPT← BREAK

****
↑ INDICATES USE OF CTROL KEY
****

Fig. 7a. Reference Card for Users of Trainer, p. 1 and 4